Exhibit IND13

Magnetic Resonance Imaging, Vol. 11, pp. 95–106, 1993
Printed in the USA.

0730-725X/93 $6.00 + .00

● *Original Contribution*

MRI: STABILITY OF THREE SUPERVISED SEGMENTATION TECHNIQUES

L.P. CLARKE,* R.P. VELTHUIZEN,* S. PHUPHANICH,†
J.D. SCHELLENBERG,*[1] J.A. ARRINGTON,* AND M. SILBIGER*

*Center for Engineering and Medical Image Analysis (CEMIA) of the Colleges of Engineering and Medicine, University of South Florida, Tampa, FL 33612, USA
†H. Lee Moffitt Cancer Center and Research Institute, Department of Neurology, Tampa, FL 33612, USA

Supervised segmentation methods from three families of pattern recognition techniques were used to segment multispectral MRI data. Studied were the maximum likelihood method (MLM), *k*-nearest neighbors (*k*-NN), and a back-propagation artificial neural net (ANN). Performance was measured in terms of execution speed, and stability for the selection of training data, namely, region of interest (ROI) selection, and interslice and interpatient classifications. MLM proved to have the smallest execution times, but demonstrated the least stability. *k*-NN showed the best stability for training data selection. To evaluate the segmentation techniques, multispectral images were used of normal volunteers and patients with gliomas, the latter with and without MR contrast material. All measures applied indicated that *k*-NN provides the best results.

Keywords: **MRI; Image segmentation; Pattern recognition methods; Neural networks.**

INTRODUCTION

MRI segmentation methods provide a quantitative basis for enhancing image contrast between normal tissues and determining tissue volumes as diagnostic indicators of disease processes.[1–3] Similarly, segmentation techniques provide methods to improve the boundary definition between pathology and edema or tissues that have similar MR relaxation parameters.[4–8] Their application to three dimensional (3D) MR image data sets should allow improved diagnosis by display with removal of normal tissue surrounding pathology[9–11] and should prove useful for 3D conformal radiation therapy treatment planning and electronic surgery simulation.[9,12] With the introduction of rapid 3D MR imaging methods, segmentation techniques may provide a basis for intelligent image fusion of multiple 3D data sets, provided corrections for image registration are applied.[9,12,13]

MR imaging systems measure spatial distributions of several distinct parameters, such as the tissue relaxation and proton density parameters. Analogous to the data sets generated by, for example, the NASA Land-Sat, which are processed in much the same way as will be described, the MR measurements are called multispectral data sets. The relative pixel intensities in the multispectral data set for each tissue class result in the formation of related tissue clusters in feature space. If image intensity based methods are used, and a large number of images (N) are generated for each anatomical slice using different radiofrequency (RF) pulse sequences, a higher order (N) feature space is obtained that further improves segmentation, not possible by visual evaluation of individual images in two dimensions. Pattern recognition techniques can be used to resolve the signal magnitude distributions as displayed in the different MR images into a probability map of tissue types, and hence can inherently provide a greater confidence level of image interpretation than simplistic gray scale approaches used for X-ray CT or single MR images.[10,14–16]

Many pattern recognition methods used for image segmentation require training data sets to identify each tissue type and are referred to as *supervised segmen-*

RECEIVED 3/31/92; ACCEPTED 6/29/92.
Address correspondence to Laurence P. Clarke, PhD, Associate Professor of Radiology, Physics and Computer Science, and Director, Center for Engineering and Medical Image Analysis (CEMIA), College of Medicine, University of South Florida, 12901 N. Bruce B. Downs Blvd., MDC Box 17, Tampa, FL 33612-4799.
[1]Current address: Picker International, Cleveland, OH.

tation techniques.[17,18] The stability of segmentation therefore depends on the accuracy of locating a region of interest (ROI) for training. Training ROIs may be selected within the same image slice, adjacent slice or slice from a different imaging session or patient. Hence the stability of segmentation for interslice, interstudy, and interpatient (training and classification) is important to evaluate for practical application of these methods.[5,19] The influence of these factors was studied for the proposed supervised segmentation methods.

In this paper methods from three families of supervised pattern recognition techniques for tissue classification are compared. The maximum likelihood method (MLM) is a representative for the parametric methods;[17] k-nearest neighbors (k-NN) is an exemplar from the family of non-parametric statistical methods;[17,18] and finally, a backpropagation artificial neural net (ANN) is a nonstatistical method from the artificial intelligence techniques.[5,20] The proposed image intensity based methods were evaluated by imaging normal volunteers as well as patients with gliomas, generating spin echo (SE), T_1- (longitudinal), T_2- (transversal), and proton density weighted images. The patients were scanned before and after MR contrast material (0.1 mmol/kg gadolinium-DTPA), the latter used as ground truth for the tumor boundary definition.[21-23]

MATERIALS AND METHODS

Pattern recognition is a general term assigned to a large group of techniques pertaining to decision making. In the case of multispectral MR image data sets, pattern recognition refers to finding rules that allow a mapping of pixel intensity values onto different tissue types. When MR image intensities for a particular anatomical location (T_1-, T_2-, and proton density weighted images) are considered, then these values may allow us to decide that a pixel (*picture element*) contains a particular tissue type. The decision will be based on the prior knowledge of the above intensity values, determined by the selected training ROIs and how close the training data sets are compared to the measured pixel intensity values. The measured pixel distribution for each tissue class is often assumed to be Gaussian, such as for the MLM method used in this work, which would allow the use of a simple function to determine the probability that a set of values belonging to a pixel represent a particular tissue type.[5]

All the possible combinations of the input values that would lead to a decision for a particular tissue type form a cluster in what is referred to as the *feature space*. This is illustrated in Fig. 1. These clusters can be very distinct, as is the cluster for cerebral spinal fluid (CSF), where tissue relaxation parameters are well differentiated. Depending on the values available for the decision, these clusters can overlap. For example, when only proton density and T_2-weighted images were available, white and gray matter cannot be fully separated. This point is illustrated by the two dimensional (2D) plots of 3D feature space shown in Fig. 1, where several tissue clusters overlap in each 2D projection displayed. Because there is a third measurement parameter ($N = 3$), a decision can be made in 3D feature space. Here lies one of the advantages of pattern recognition on multispectral data sets over gray scale methods, particularly if multiple image sets $N \geq 3$ are acquired.[17,18]

Segmentation Methods

The theoretical basis for the proposed segmentation methods has been extensively reviewed in the literature.[1,4,24-27] This work will focus on the clinical evaluation of the stability of the segmentation methods using current spin echo imaging protocols.

The Maximum Likelihood Method (MLM). This classifier is based on Bayes decision rule which maximizes the a posteriori probabilities. This rule can be written as:

$$X \text{ is in class 2 if } P_1 p_1(X) < P_2 p_2(X) \ . \qquad (1)$$

X is a feature-vector, in this case a vector with as elements the intensities for a pixel in each of the MR images, capital P stands for a priori probability for a class, and the lower case p is the conditional probability for X given that it is in that class. The conditional probabilities are given by the multivariate normal density function:

$$p_i(X) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\{-\tfrac{1}{2} d^2\} \ , \qquad (2)$$

where n is the dimension of feature vector X and d is a distance measure (the Mahalanobis distance) described by:

$$d^2 = (X - M)^T \Sigma^{-1} (X - M) \ , \qquad (3)$$

with M the mean vector and Σ the covariance matrix. Both M and Σ are estimated from the training set. The a priori probabilities are assumed to be equal. The algorithm first uses the training samples to calculate the mean vector and covariance matrix. Then for each pixel the probabilities for membership of each class is calculated using Eqs. (3) and (2). Finally a decision is made for the classification of each pixel using Eq. (1), choosing the class with the highest probability (Max-

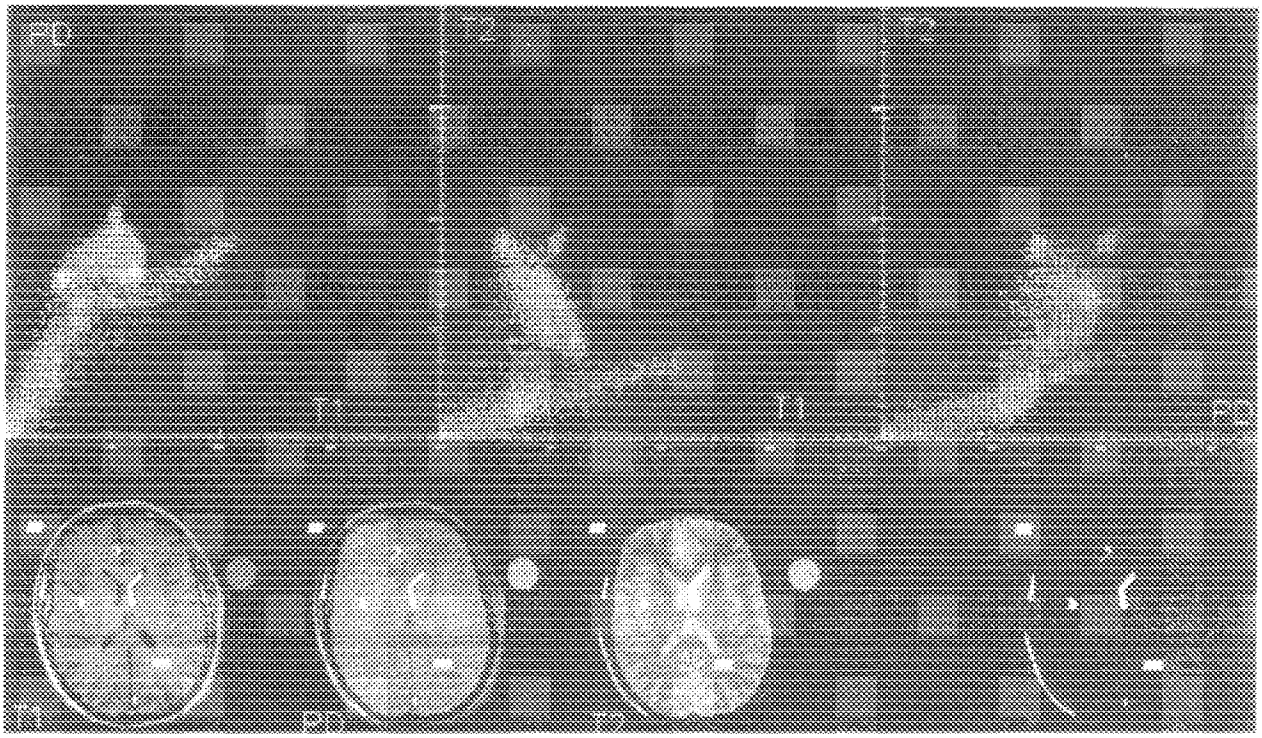

Fig. 1. Display used for selection of training data. Top graphs represent 2D scatterplots of the three possible combinations of two images; on the bottom row the T_1-, proton density, and T_2-weighted images are shown with the selected regions overlaid. The intensity values of the regions are overlaid in the scatterplots. In the lower right the regions of interest (ROIs) are shown separately. The regions are shown in color on the computer screen to the operator.

imum Likelihood). The method is described in more detail elsewhere,[17] and is used in multispectral MR image analysis.[7]

k-*Nearest Neighbors* (k-*NN*). This method does not use any particular distribution, but estimates the a posteriori probabilities from the distribution of neighbors in the feature space. *k*-NN calculates the distance from the pixel feature vector to all the training pixel feature vectors and calculates an estimate of the a posteriori probabilities from the frequency of the labels (classes) of the *k* nearest neighbors, where *k* is an odd integer.[17,18] *k*-NN is a member of the same family of nonparametric classifiers as Parzen windows, which method has been used by other researchers.[28] The difference with the Parzen windows approach is that the window size for *k*-NN is adjusted automatically to include only *k* neighbors.[18] The implementation of the algorithm as described is the *voting k*-NN, the value for *k* used was 7, and the choice was based on the fact that increase of *k* did not significantly change the segmentations.

Artificial Neural Network (ANN): An ANN is a massively parallel architecture made up of a number of units, which are linked by some connections. These units are capable of accepting incoming signals, performing some boolean or arithmetic processing on the data and sending signals out over some or all of the connections. The storage of "knowledge" is accomplished by altering the weight associated with each connection.[20] The behavior of the nodes in the used neural net is described by:

$$\text{net}_i = f(\Sigma_j w_{ij} I_j) \tag{4}$$

with

$$f(x) = [1 + \exp(-x)]^{-1} \tag{5}$$

$f(x)$ is called the *activation* function; w_{ij} is the weight for the connection between input j, I_j, and node number i with output signal net_i.

The architecture used for the segmentation in this work is the traditional feed forward artificial neural net with backpropagation training, customarily called *backprop*. The labelled training data are offered to the input nodes of the neural network, which calculates an output. The output is compared to the desired output (the label assigned to the training vector), and used to

calculate an update of the weights in the neural network to minimize the error. This is done with each of the labelled training data, and repeatedly so until the total error is smaller than a pre-set threshold. In this work the segmentations are obtained using backprop with two hidden layers as previously reported by these investigators.[5,6] The number of neurons in each layer depends on the number of outputs but is generally in the order of 14 nodes in the first hidden layer and 8 hidden nodes in the second hidden layer.

Selection of Training Regions

The method developed for the selection of training ROIs for supervised image segmentation is shown in Fig. 1. In this work MRI images of the brain of normal volunteers were acquired using a spin-echo (SE), multislice 2DFT technique (5mm thickness, 256 × 256 matrix, 2.5 mm interslice gap), namely a T_1-, T_2-, and spin-density weighted image, as shown in Fig. 1 for a single axial slice. Two dimensional projections of the 3D feature space with the feature vectors of the selected ROIs highlighted are also plotted in Fig. 1. Normally the ROIs for each tissue class are color coded (not shown) for visual inspection. The 2D displays of feature space are fairly standard in the literature.[3,25,28] The six tissue types chosen for training ROIs in Fig. 1 were white matter; gray matter; cerebrospinal fluid; muscle; fat; and a class for the background (air and bone). For patient studies, additional ROIs would be chosen for tumor (gadolinium enhancement) and edema. The operator interface allowed the user a concurrent view of the images and the feature space as shown in Fig. 1. It was possible to select or modify training data directly in the spatial domain or in the feature space, thus minimizing errors in the selection of the training sets.[25] The software was developed for interactive selection of ROIs using the Interactive Data Language (Research Systems, Inc., Boulder, CO).[29]

Imaging Protocols

The training and classification was done on comparable studies of three normal male volunteers in the age group 25–45 yr. The images were taken with a 1.5-T system, using a quadrature RF head coil with a uniform RF field in the transverse plane.[30] Three spectral images were taken in the transverse plane in the center of the coil, with the following spin echo sequences: (TR = 600 msec, TE = 20 msec); and (TR = 3000 msec, TE = 20 msec, TE = 80 msec). All scans were taken with a small gel phantom placed close to the head of the patient, within the RF coil, as a basis of image intensity standardization for interslice, interscan or interpatient analysis.[5,30] Patients were imaged using the same pulse sequence as the normal volunteers. Initial measurements on volunteers used single slice 2DFT techniques (idealized data); later volunteer studies and patient protocols used multislice techniques (50% gap). Standard quality control measurements were performed prior to each study, including uniformity and signal-to-noise ratio measurements.[30] Patients were imaged with the multispectral data set prior to the use of MR contrast material and after injection, without removal of subject from the MR table. The error in table movement was within ± 1mm.

Evaluation Methods

Three issues were addressed in comparing the different methods: execution speed, stability of segmentation and correctness of classification. Execution speed was determined using a representative data set/training set. Stability was measured for different sources of the training data. First, the stability for choice of the training regions (ROIs) was measured. An operator picked the regions in five different locations. Each of these locations were correct but on a different anatomical site, thus simulating the effect of inter- and intraoperator variability. The results of the classification were calculated and displayed as a set of invariant pixels and a display of pixels of which the classifications depended on the particular choice of training regions. Secondly, the stability for choice of the source image for the training data was measured in three categories: interslice (another slice in the same multislice study), interscan (same location in the subject, but from an MR image acquired on a different date) and interpatient (training data obtained from a data set of a different subject) classifications. The procedure to obtain a measure for the stability is outlined here for the case of interslice classifications.

Consider a number of slices, say six, in a multislice dataset. In each of those slices regions of interest are marked, giving six different sets of training data. Then one of the slices can be classified using each of those sets of training data, thus yielding six different classifications. Some pixels will be assigned to the same class in every one of those six results, but there will also be pixels that are sometimes classified differently. The percent varying pixels is then the ratio of the number of pixels not all six times in the same class and the total number of pixels in the image. Similarly, the resulting classifications can be compared with the ROIs that were picked on the specific slice under consideration. If in any of the six segmentations a pixel is not classified in the same class as the operator had picked, that pixel is counted as a misclassified ROI pixel. The percentage is then the ratio of the total number of misclassified ROI pixels and the number of pixels assigned to an ROI by the operator.

For tumor cases for which pre- and postcontrast images were available, the pixels representing tumor were determined using a ratio image. The histogram of a region around the enhancement was used to establish the intensity cut-off point of the enhancement for both the pre- and postcontrast T_1-weighted images. The tumor was then found as the pixels with the precontrast intensities below the threshold, the postcontrast intensities above the threshold and showed an increase in intensity of at least 50%.

Correctness of segmentation was evaluated by a team of three experienced MR radiologists by comparison of the segmentation with the MR films. The radiologists were blinded to the patient history and to each others responses. A questionnaire allowed a scoring system in which the radiologists could grade the performance of the methods on correctness of differentiation between tissues or groups of tissues. Each radiologist gave a score between 0 (totally wrong) and 10 (differentiation between tissues exactly correct). Although this method does not allow estimation of absolute correctness (ROC-analysis), given the fact that ground truth is not available this method gives an indication of the relative performance in terms of correctness of the classifiers.

The computer hardware used was a Sun 4/470 server with 32 MByte and a GX framebuffer. All the programs were written in a high level data analysis language IDL[29] that permitted a highly efficient interactive analysis of the data and code development.

RESULTS

Visual Comparison of Segmentation Methods

Representative results of the three segmentation methods for a 45-yr-old male volunteer are shown in Fig. 2. The difference in segmentation is small for those tissues that are well differentiated due to tissue-

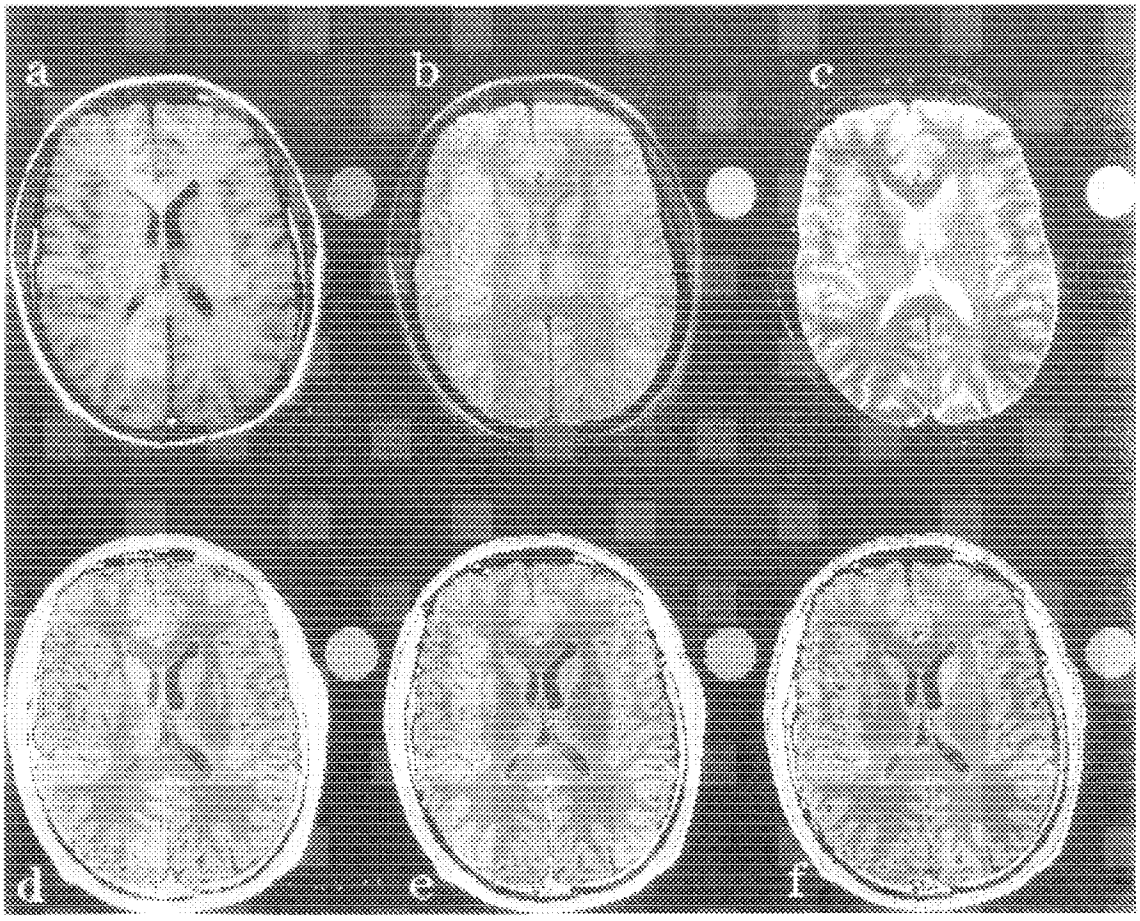

Fig. 2. Representative results for the three segmentation methods. Top row: MR images. (a) T_1-weighted image (TR = 600 msec, TE = 20 msec); (b) Proton density weighted image (TR = 3000 msec, TE = 22 msec); (c) T_2-weighted image (TR = 3000 msec, TE = 80 msec). Bottom row, segmentation results: (d) Maximum likelihood method; (e) k nearest neighbors, $k = 7$; (f) feedforward backpropagation artificial neural net.

specific MR relaxation or proton density related parameters. However, for tissues that are not well differentiated such as white and gray matter, differences in segmentation were observed particularly between the MLM and the *k*-NN and ANN technique. This is partly attributed to the assumption for the MLM method, that the tissue parameter distribution in Gaussian, which may often not be the case as previously shown by these investigators.[5,19] Analysis of the distributions of the tissue clusters has revealed that for some tissues, in particular white matter and pathology, the assumption that the intensities in these clusters are arranged according to the multivariate Gaussian distribution is false.[5,19]

Results for the patient studies are shown in Figs. 3 and 4. It can be observed, as will be quantified in the next sections, that the results for MLM suffer from noise and that the ANN and *k*-NN results are quite comparable.

Stability Dependence on Training ROI

The stability of segmentation for different training ROI's usual normal volunteer studies is illustrated in Fig. 5. After careful visual inspection of the images of the *k*-NN method appeared to be the most stable. This observation is also represented in Fig. 6 where bar graphs illustrate the relative stability of each tissue class for these segmentations. The quantitative analysis again showed that the *k*-NN method was the most stable. More specifically the *k*-NN method shows the smallest differences between the number of pixels that are always classified as a particular tissue type and the number of pixels assigned at least once to this class, represented by the total height of the bar graph.

Interslice, Scan, and Patient Stability of Segmentation

The interslice, interscan and intervolunteer stability for selection of training ROIs and segmentation for normal volunteers is shown in Figs. 7 and 8. The percentages were obtained as described in section 2.6. The results displayed represent a study obtained from 11 multispectral MR single slice data sets acquired on three volunteers, and a total of 174 different segmentations. Comparison with results for multislice acquisitions showed no significant differences. It can be

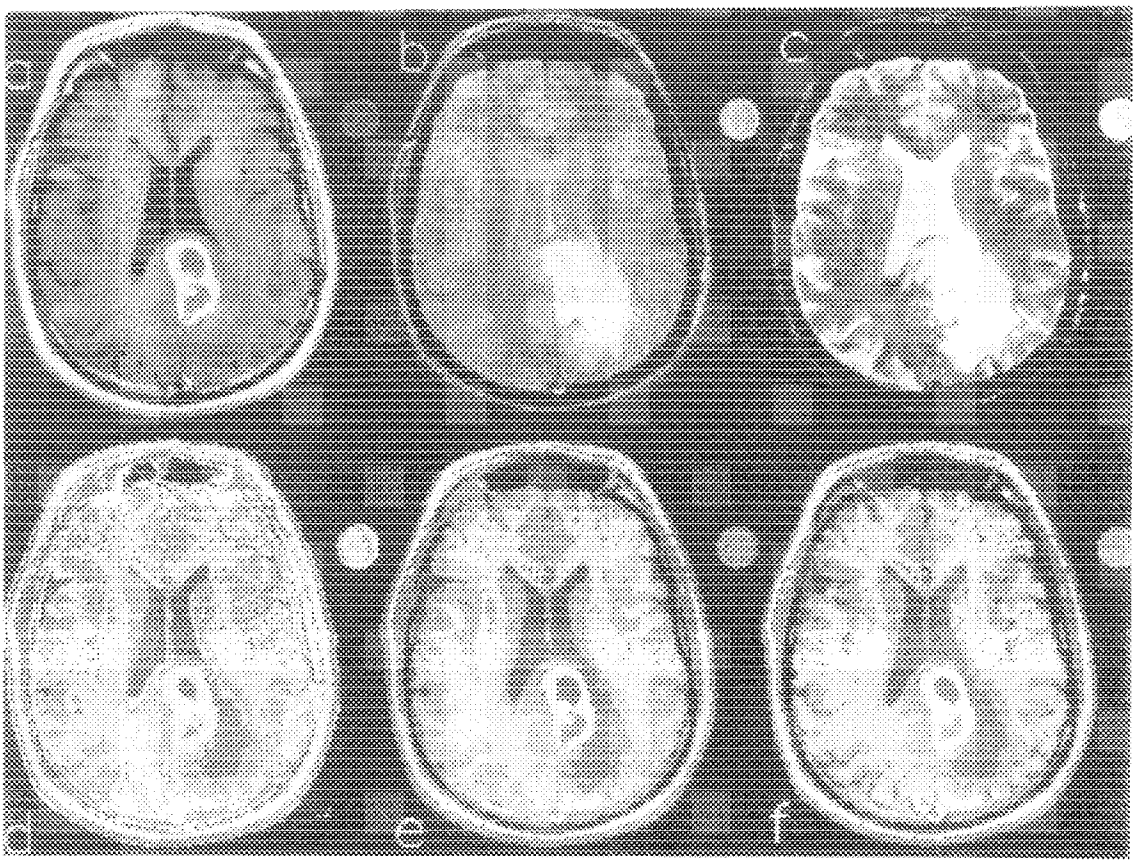

Fig. 3. Segmentation results for a 71-year-old male patient study. (a) T_1-, (b) proton density, and (c) T_2-weighted images. Segmentation results: (d) MLM, (e) *k*-NN, and (f) ANN.

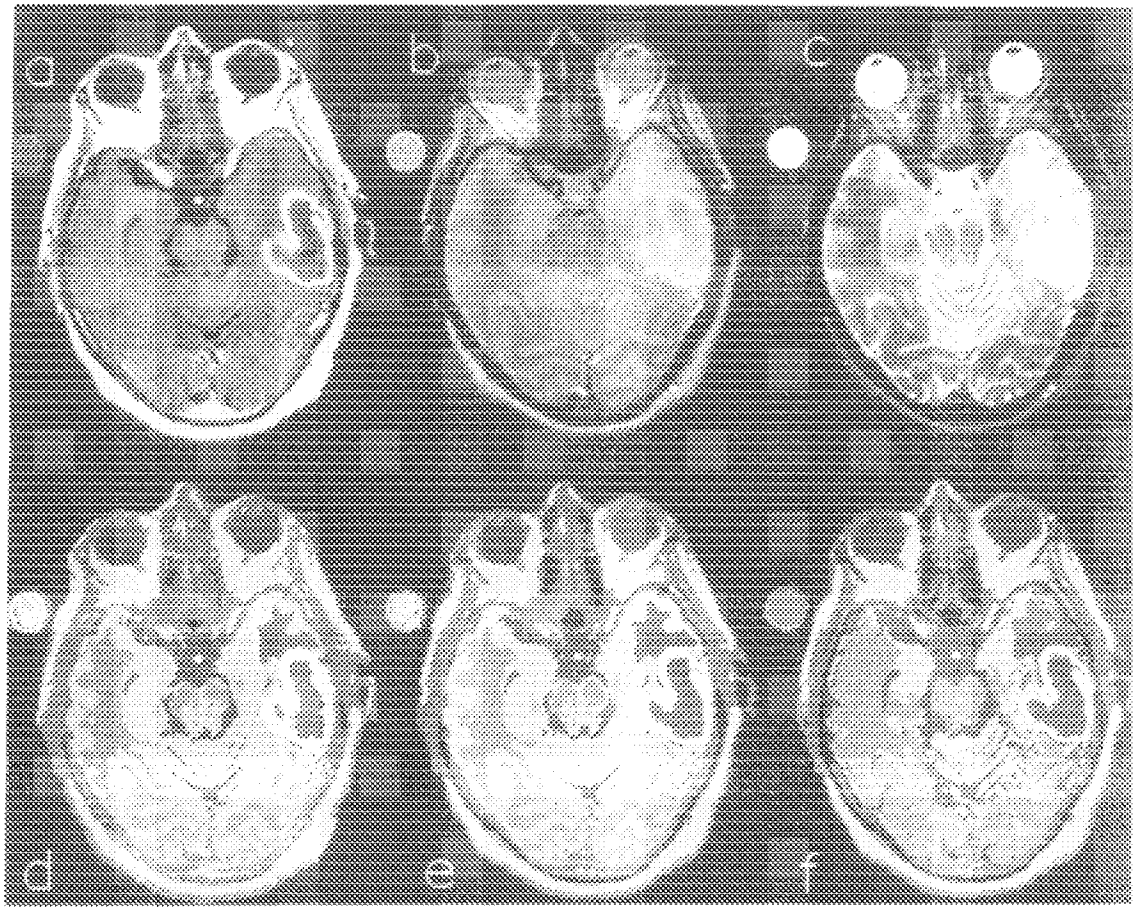

Fig. 4. Segmentation results for a 63-year-old female patient study. Labels as in Fig. 3.

seen that k-NN shows superior stability for all criteria applied, and MLM gives the worst results.

It should be noted that the stabilities must be compared between the different methods and not between the different criteria. The influence of the operator is very large, as can be seen in the top graphs (ROI dependence). Within one criterion, the same regions of interest are used for each method and thus results can be compared to each other. Between criteria different images and different training regions are used and results are not comparable.

Clinical Evaluation and Interobserver Interpatient Analysis

Radiologists. Correctness of classification was evaluated by a team of three experienced radiologists who were blinded to the patient history and to each others' responses. Evaluated were four studies of normals and two patient studies. A score between 0 (totally incorrect differentiation) and 10 (differentiation between tissues exactly correct) for each method was obtained for the classification of the intracranial part of the images. The questionnaire asked to take into consideration "the sharp margination between gray and white matter and cerebrospinal fluid as well as the identification of known anatomic parameters, i.e., using standard MRI parameters of tumor definition (tumor extend, margins, necrosis, edema)," in other words using MR contrast enhancement.[22,23] The results of the evaluation are tabulated in Tables 1 and 2. Although disagreement between the radiologist existed on points, there was consensus on the overall ranking of the segmentation methods.

The k-NN method scored the best for correctness of classification, particularly in differentiating between tumor and edema and between gray and white matter as previously observed for the normal volunteer study, that is, where the relaxation behavior between the tissues is frequently not well differentiated. The MLM method performed comparatively poorly, particularly in differentiating between normal tissues and pathol-

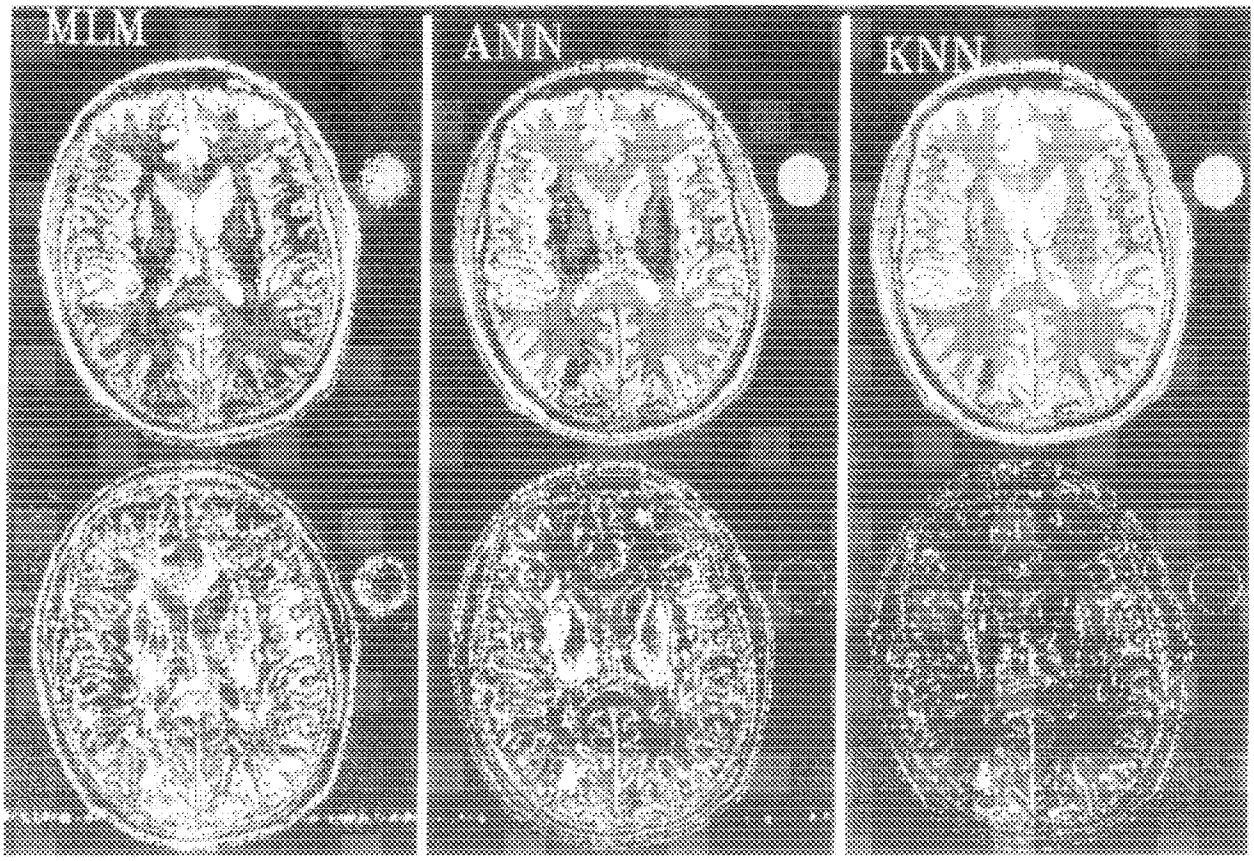

Fig. 5. Classification results for five different sets of training regions. Top row: pixels that are assigned the same class for each of the sets of training ROIs. Bottom row: varying pixels are displayed in white.

ogy. The ANN method was better than MLM but did not perform as well as *k*-NN for gray and white matter and tumor and edema differentiation.

Enhancement. The correctness of classification was also evaluated by using the postcontrast images as ground truth, with a knowledge of the influence of tumor staging and blood–brain barrier on the spatial and temporal changes of the contrast material.[21-23] The images and segmentation results are shown in Figs. 3 and 4. In this quantitative evaluation, the correct tumor boundary prior to the use of contrast using the three segmentation methods, was compared to the MR contrast study. The tumor boundary was defined by thresholding the ratio image of the pre- and postcontrast T_1-weighted image as described above. The results obtained by this method are tabulated in Table 3, and appears to confirm the previous observations, that the *k*-NN method provided the best clinical definition of the tumor extent.

Execution Times for Segmentation

The execution time for each segmentation for one multispectral data set is shown in Table 4, when the Sun 4/470 is used without the use of parallel processing hardware. The MLM was the faster method, the *k*-NN taking the longest time. The training period for the MLM method is the time needed to calculate the mean vectors and covariance matrixes for the training data. *k*-NN does not have a training time; for each pixel the same calculations have to be done and no "training" before per-pixel classification is started is done.

Table 1. Evaluation of image segmentation methods by experienced radiologists in percentage of the perfect score. Errors indicated are errors in the mean

Differentiation between tissues	MLM	*k*-NN	ANN
CSF and brain parenchyma	74 ± 2	81 ± 3	82 ± 3
Gray matter and white matter	74 ± 3	83 ± 2	76 ± 4
Normal tissues and pathology	68 ± 5	85 ± 5	83 ± 2
Tumor and edema	73 ± 6	92 ± 2	78 ± 6
Total	73 ± 2	83 ± 1	79 ± 2

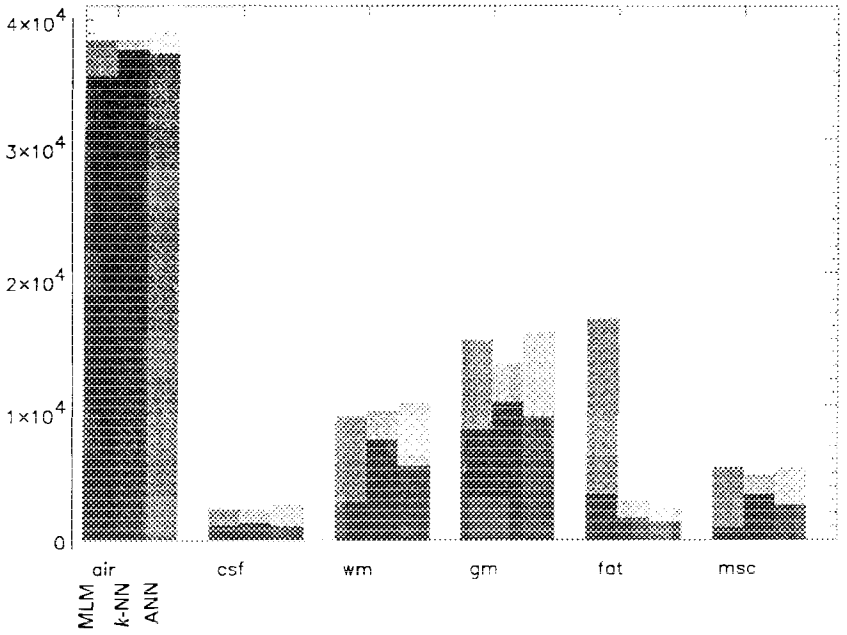

Fig. 6. Stability per tissue class for ROI selection as follows from Fig. 2. Columns from left to right: MLM, *k*-NN, ANN. Lower part of the columns: number of pixels that are assigned the same class for all choices of ROI. Whole column: number of pixels classified at least once in that class.

DISCUSSION

In recent years several groups have attempted construction of tissue maps using pattern recognition techniques on data sets acquired with sequences used in the current daily practice of diagnostic imaging, in particular with standard spin-echo sequences[4,25,26,31] or with fast imaging techniques.[24] Some authors have described very simple derivations of pattern recognition techniques, using a human operator to define clusters in feature space[32] or some simple heuristic operating in feature space[3] with limited results.

Confidence in the correctness of the segmentation results is affected by the stability of the method. The several supervised segmentation methods proposed in the literature[3,4,10,25,26,32] do not address the issue of stability of segmentation. This work focuses on the problem of the stability of supervised image segmentation techniques and related issues of image interpretation for both normals and patients with selected brain tumors. In this preliminary study it is shown that the results for the maximum likelihood method heavily depend on the training data. Also, the MLM results are anatomically the least correct, probably related to insufficient knowledge of the distribution of pixel intensities. In a recently published reliability study,[33] five pattern recognition methods were tested for multispectral MR image segmentation, of which only MLM was in common with the study presented here. The authors found MLM to be the best classifier of the methods tested. It should be noted that in that report error rates of 50% for the "best" classifier MLM were observed.

Table 2. Evaluation of segmentation methods by patient type in percent of maximum score. Errors indicated are errors in the mean

Patient type	MLM	*k*-NN	ANN
Normals	79 ± 2	83 ± 2	83 ± 3
Patients	68 ± 3	84 ± 2	75 ± 3

Table 3. Percent false negatives and percent false positives in percent of the enhancement area

	MLM	*k*-NN	ANN
Patient 2			
% false pos	15	18	17
% false neg	16	9	18
Patient 4			
% false pos	223	114	54
% false neg	10	14	43

Table 4. Execution times for segmentation of one data set in seconds. Sun 4/470 GX, 32MB; IDL

	Training	Segmentation
MLM	0.06	7.7
k-NN	–	213
ANN	60.0	122

The backpropagation ANN showed much better performance, both in terms of stability and correctness. The ANN method does not require assumptions on the distribution of pixel intensities, and hence the segmentations may reflect more the true distributions of the dataset. The fixed architecture and the specific choice of the number of nodes and connections may have influenced the results. We are currently investi gating the stability of segmentation using the cascade correlation method,[8] which would further reduce the number of operator decisions necessary. Application of *k*-NN requires only one parameter to be set by the operator: *k*, the number of nearest neighbors used to estimate the a posteriori probabilities. In this limited study, *k*-NN performed much better than MLM and slightly better than ANN. These techniques are now being applied to 3D data sets to determine the stability over the full volume image by using each individual slice for training and classification. The 3D multispectral methods are useful for improved quantitative measurement of tumor volume and related response measurements for radiation or chemotherapy, and should provide better tumor boundary definition compared to simple gray scale approaches that do not enhance pathology.[10,15,16]

The work reported here was focussed on performing a relative comparison of supervised segmentation methods. Other factors affect the stability of segmentation such as patient and instrument related factors. Patient related factors include the influence of age, sex, or prior tumor therapy schemes on MR relaxation parameters and related MR image contrast between tissues within the multispectral data set. Pattern recognition methods that train and classify on the same data set should partly minimize the instability of segmentation. Instrument related factors include RF field inhomogeneity, slice profile, and related partial volume effects and problems related to image registration. RF field inhomogeneity can be minimized using multielement resonator coils and can be corrected for using tailored RF sequences or using empirically derived correction matrices.[30] Registration correction for rigid body motion is possible.[13] However, advances in MR technology may reduce many of the instrument related factors that give rise to instability of segmentation such as rapid 2D or 3DFT imaging methods with reduced effective slice thickness and various methods for motion and flow compensation.

Alternative supervised segmentation methods are possible as reported elsewhere. Of particular interest is the use of artificial neural network techniques with dynamic architecture as opposed to a fixed architecture used here. We are currently exploring cascade cor-

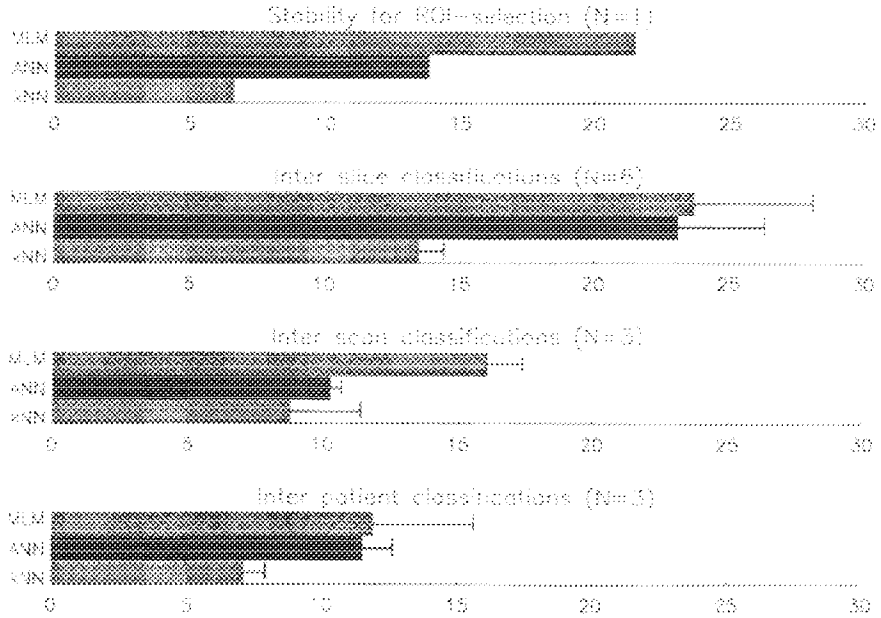

Fig. 7. Stability of segmentation methods plotted as percent pixels not always assigned the same class.

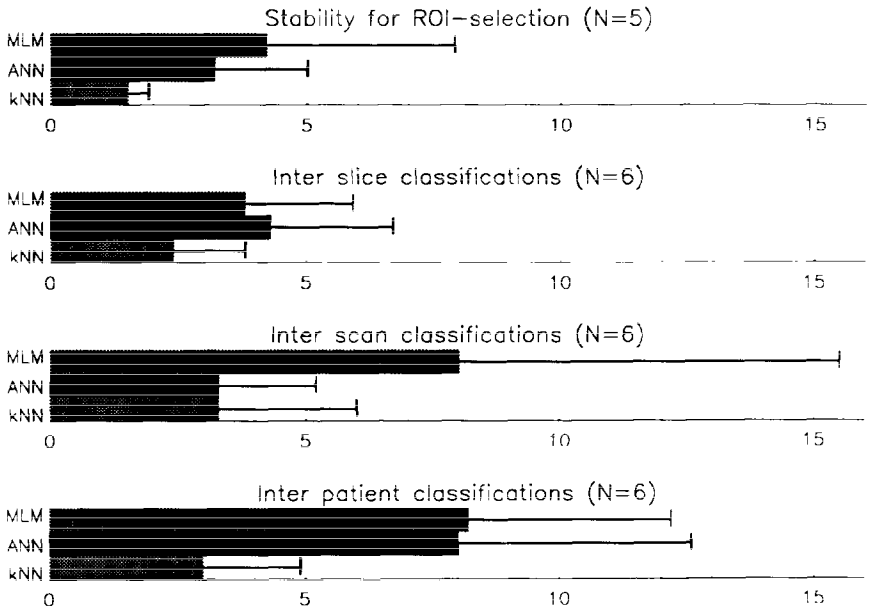

Fig. 8. Stability of segmentation methods plotted as percent pixels assigned a class by the operator that were classified differently in at least one instance of training data variation.

relation techniques[8] that may show greater stability with different training data. Similarly, unsupervised segmentation methods are possible, such as fuzzy clustering techniques [Fuzzy c-means (FCM)][34] or alternatively partially supervised FCM, that allow better differentiation of tissue clusters in feature space than FCM as reported by these investigators.[8,35]

CONCLUSION

In this work it is shown that computer analysis can provide improved boundary definition between tissues of similar MR relaxation parameters such as tumor and edema and some confidence level in the interpolation of these boundaries as shown in Figs. 3 and 4. The stability of three well developed pattern recognition techniques for image segmentation was investigated, showing that ANN and *k*-NN perform considerably better than MLM, ensuring reproducibility of segmentation results. Further increased reliability is expected from newer methods as self-organizing artificial neural nets and fuzzy c-means methods.

We are actively exploring these supervised and unsupervised segmentation methods to remain in phase with the improvements and speed of MRI technology. The use of digital RF systems, the progress in multielement quadrature coil designs and the development of RF pulse sequences that are not MR inherent parameter based (**k**-space trajectory techniques, rapid acquisition with relaxation enhancement, fast magnetization preparation techniques, and proton perfusion and diffusion) provide either improved image signal/noise ratio and increase the *N* dimensional feature space and should inherently result in improved image segmentation. Pattern recognition methods should therefore play an increasingly important role for image fusion of multispectral 3D data sets and provide an improved confidence level for image interpretation using a 3D workstation for MRI data analysis.

Acknowledgments- This investigation was supported in part by grants from NASA, American Cancer Society/Florida Division, Siemens Medical Systems, New Jersey and Sun Microsystems, California.

REFERENCES

1. Hyman, T.J.; Kurland, R.J.; Levy, G.C.; Shoop, J.D. Characterization of normal brain tissue using seven calculated MRI parameters and a statistical analysis system. *Magn. Reson. Med.* 11:22–34; 1989.
2. Rajanayagam, V.; Fabry, M.; Gore, J.C. In-vivo quantitation of water content in muscle tissue by NMR imaging. *Magn. Reson. Imaging* 9:621–625; 1991.
3. Kohn, M.I.; Tanna, N.K.; Herman, G.T.; Resnick, S.M.; Mozley, P.D.; Gur, R.E.; Alavi, A.; Zimmerman, R.A.; Gur, R.C. Analysis of brain and cerebrospinal fluid volumes with MR imaging. *Radiology* 178(1):115–122; 1991.
4. O'Donnell, Mm; Gore, J.C.; Adams, W.J. Toward an automated analysis system for nuclear magnetic resonance imaging. II. Initial segmentation algorithm. *Med. Phys.* 13(3):293–297; 1986.
5. Schellenberg, J.D.; Naylor, W.C.; Clarke, L.P. Appli-

cation of artificial neural networks for tissue classification from multispectral magnetic resonance images of the head. Proceedings of the Third Annual IEEE Symposium on Computer-Based Medical Systems. Chapel Hill, NC; June 3–6, 1990.

6. Clarke, L.P. Comparison of Bayesian maximum likelihood (MLM) and artificial neural network for supervised tissue classification and 3D segmentation by MRI. 33rd Annual Scientific Symposium of American Association of Physicists in Medicine (AAPM). San Francisco; July 1991; *Med. Phys.* 18(3):673; 1991.
7. Vannier, M.W.; Butterfield, R.L.; Jordan, D.; Murphy, W.; Levitt, R.; Gado, M. Multispectral analysis of magnetic resonance images. *Radiology* 154:221–224; 1985.
8. Bensaid, A.M.; Hall, L.O.; Clarke, L.P.; Velthuizen, R.P. MRI Segmentation using supervised and unsupervised methods. Proceedings of the 13th Annual IEEE Eng Med & Biol. Orlando, FL; Oct 31–Nov 3, 1991.
9. Levin, D.N.; Hu, X.; Tan, K.K.; Galhotra, S.; Pelizzari, C. The brain: integrated three-dimensional display of MR and PET images. *Radiology* 172:783–789; 1989.
10. Xiaoping, H.O.; Alperin, N.; Levin, D.N. Visualization of MR angiographic data with segmentation and volume rendering techniques. *J. Mag. Res. Imag.* 1(5):539; 1991.
11. Bomans, M.; Hohne, R.H.; Laub, G. Improvement of 3D acquisition and visualization by MRI. *Magn. Reson. Imaging* 9:579–609; 1991.
12. Tan, K.K.; Levin, D.N.; Pelizzari, C.A.; Chen, G.T.Y. Interactive stereotactic localization of brain anatomy. *Radiology* 177(P):217–351; 1990.
13. Alpert, N.M.; Bradshaw, J.F.; Kennedy, D.; Correia, J.A. The principal axes transformation—a method for image registration. *J. Nucl. Med.* 31:1717–1722; 1990.
14. Brummer, M.E.; Mersereau, R.M.; Eisner, R.L.; Lewine, R.R.J. Automatic detection of brain contours in MRI datasets. Proceedings of the 12th International Conference Information Processing in Medical Imaging. Wye, UK; July 7–12, 1991; Berlin: Springer Verlag; 1991.
15. Menhardt, W.; Schmidt, K.H. Computer vision on magnetic resonance images. *Pattern Recog. Lett.* 8:73–85; 1988.
16. Suzuki, H.; Toriwaka, J. Automatic segmentation of head MRI images by knowledge guided thresholding. *Comput. Med. Imaging Graph.* 15(4):233–240; 1991.
17. Duda, R.; Hart, P. Pattern Classification and Scene Analysis. New York: Wiley; 1973.
18. Fukunaga, K. Introduction to Statistical Pattern Recognition. 2nd ed. New York: Academic Press; 1990.
19. Velthuizen, R.P.; Clarke, L.P.; Hall, L.O.; Bensaid, A.M. Multispectral 3D MRI segmentation using knowledge based systems. *Med. Phys.* 18(3):622; 1991.
20. Fahlman, S.; Hinton, G. Connectionist architectures for artificial intelligence. *IEEE Computer* 20(1):100–109; 1987.
21. Russel, E.J. Imaging of intracranial neoplasms. *Curr. Opin. Radiol.* 2:4–17; 1990.
22. Schorner, W.; Lanaido, M.; Kornmesser, W.; Felix, R. Comparison of multiecho and contrast enhanced MR scans: Image contrast and delineation of intracranial tumors. *Neuroradiology* 31:140–147; 1989.
23. Johnson, P.C.; Hunt, S.J.; DFrayer, B.P. Human cerebral gliomas: Correlation of post mortem MR imaging and neuropathologic findings. *Radiology* 170:211–217; 1989.
24. Koenig, H.A.; Bachus, R. Tissue discrimination in magnetic resonance 3D data sets. Poster: Sixth Annual Meeting of the Society of Magnetic Resonance in Medicine. New York: August 1987.
25. Vannier, M.W.; Speidel, C.M.; Rickman, D.L. Magnetic resonance imaging multispectral tissue classification. *News in Physiol. Sci.* 3:148–154; 1988.
26. Jungke, M.; von Seelen, W.; Bielke, G.; Meindl, S.; Krone, G.; Grigat, M.; Higer, P.; Pfannenstiel, P. Information processing in nuclear magnetic resonance imaging. *Magn. Reson. Imaging* 6:683–693; 1988.
27. Just, M.; Thelen, M. Tissue characterization with T1, T2, and proton density values: Results in 160 patients with brain tumors. *Radiology* 169:779–785; 1988.
28. Gerig, G.; Martin, J.; Kikinis, R.; Kübler, O.; Shenton, M.; Jolesz, F.A. Automating segmentation of dual-echo MR head data. Information Processing in Medical Imaging, Proceedings of 12th International Conference Information Processing in Medical Imaging. Wye, UK; July 7–12, 1991; Berlin: Springer-Verlag: 175–187; 1991.
29. Interactive Data Language (IDL) Version 2.1, Users Guide. Research systems, Inc. Boulder, CO.; April 1991.
30. Glennon, D.T. Correction of magnetic resonance image non-uniformity. University of South Florida; December 1991. Masters Thesis.
31. Merickel, M.B.; Carman, C.S.; Brookeman, J.R.; Ayers, C.R. Image analysis and quantification of atherosclerosis using MRI. *Comput. Med. Imaging Graph.* 15(4):207–216; 1991.
32. Windham, J.P.; Peck, D.J.; Soltanian-Zadeh, H. A vector subspace method for tissue characterization using MRI sequences. *Med. Phys.* 18(3):619; 1991.
33. Vannier, M.W.; Pilgram, T.K.; Speidel, C.M.; Neumann, L.R.; Rickman, D.L.; Schertz, L.D. Validation of magnetic resonance imaging (MRI) multispectral tissue classification. *Comput. Med. Imaging Graph.* 15(4): 217–223; 1991.
34. Bezdek, J.C. A convergence theorem for the fuzzy ISODATA clustering algorithms. *IEEE Trans on Pattern Analysis and Machine Intelligence* PAMI-2(1):1–8; 1980.
35. Hall, L.O.; Bensaid, A.M.; Clarke, L.P.; Velthuizen, R.P.; Silbiger, M.; Bezdek, J.C. A comparison of neural networks and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. *IEEE Trans on Neural Networks* (in press).